Canonical Analysis

Albers, Casper; Gower, John C.

Published in:
Journal of Classification

DOI:
10.1007/s00357-014-9146-y

Publication date:
2014

Journal of Classification 31:2-27 (2014)
DOI: 10.1007/s00357-014-9146-y

Canonical Analysis: Ranks, Ratios and Fits

Casper J. Albers

University of Groningen, The Netherlands

John C. Gower

The Open University, UK

Abstract: Measurements of p variables for n samples are collected into a $n \times p$ matrix **X,** where the samples belong to one of k groups. The group means are separated by Mahalanobis distances. CVA optimally represents the group means of **X** in an r-dimensional space. This can be done by maximizing a ratio criterion (basically one-dimensional) or, more flexibly, by minimizing a rank-constrained least-squares fitting criterion (which is not confined to being one-dimensional but depends on defining an appropriate Mahalanobis metric). In modern $n < p$ problems, where **W** is not of full rank, the ratio criterion is shown not to be coherent but the fit criterion, with an attention to associated metrics, readily generalizes. In this context we give a unified generalization of CVA, introducing two metrics, one in the range space of **W** and the other in the null space of **W**, that have links with Mahalanobis distance. This generalization is computationally efficient, since it requires only the spectral decomposition of a $n \times n$ matrix.

Keywords: Canonical analysis; Ratio form; Fit form; Mahalonobis distance; Discriminant analysis.

1. Introduction

The following is developed in the context of Canonical Variate Analysis (CVA) but, with minor adaptations, is relevant to all forms of multivariate canonical analyses. CVA was originally defined (Rao 1949) as optimizing the ratio of two quadratic forms which, as is well-known, requires the maximal eigenvalue of a two-sided eigenvalue problem. The remaining eigenvalues, and associated eigenvectors, are commonly also

Corresponding Author's Address: C.J. Albers, Department of Psychometrics and Statistical Methods, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands, e-mail: c.j.albers@rug.nl.

Published online: 29 March 2014

utilized with the justification of being optimal, conditional on orthogonality with respect to the preceding eigenvectors. Other approaches, based on optimal least-squares fits to an appropriate matrix, give a better justification for multidimensional solutions and offer a better basis for developments required such as those that arise when the number of variables exceeds the number of samples. With technological advances, this is becoming increasingly important but it brings with it questions of deficiencies in rank that do not arise in classical applications. Thus, the following discussion tries to disentangle several strands – ratio versus least-squares criteria, the role of constraints, the more-variables-than-samples problem, and the use of different metrics. Although these matters may seem to be well understood, a careful appraisal reveals that there are some subtle issues that deserve consideration. We attempt a unified approach that examines these issues in a coherent framework.

We begin by establishing notation and reviewing some well-known results (Section 2) in the classical theory, though presented from our own perspective. This Section is a necessary precursor to Section 3, the heart of the paper, on the implications for modifications required when there are more variables than samples. Section 4 gives an example.

Notation. In CVA we have measurements on each of p variables for n samples distributed among k groups of sizes $n_1 + n_2 + \dots + n_k = n$. These measurements are available in an $n \times p$ matrix $\mathbf{X}$, assumed column-centered, and therefore of rank at most $min(n-1, p)$, with group-membership given in an $n \times k$ indicator matrix $\mathbf{G}$. Here, $\mathbf{G}$ is zero except that $g_{ij} = 1$ when the ith sample belongs to the jth group. Thus $\mathbf{G1} = \mathbf{1}$ and $\mathbf{1'G} = \mathbf{1'N}$, where $\mathbf{N} = diag(n_1, n_2, \dots, n_k) = \mathbf{G'G}$.

We shall also need the idempotent matrix ${}_n\mathbf{H}_n = {}_n\mathbf{G}_k\mathbf{N}_k^{-1}\mathbf{G}'_n$ which represents an orthogonal projection from n dimensions into p dimensions. With this notation, we have the usual between and within-group orthogonal decomposition:

$$ {}_n\mathbf{X}_p = {}_n\mathbf{G}_k\mathbf{N}_k^{-1}\mathbf{G}'_n\mathbf{X}_p + [\mathbf{I} - {}_n\mathbf{G}_k\mathbf{N}_k^{-1}\mathbf{G}'_n]_n\mathbf{X}_p = \mathbf{HX} + (\mathbf{I}-\mathbf{H})\mathbf{X}, \quad (1) $$

with associated orthogonal analysis of variance that the Total sum-of-squares ($\mathbf{T}$) is the sum of the Between-Group sum-of-squares ($\mathbf{B}$) and the Within-Group sum-of-squares ($\mathbf{W}$):

$$ \begin{array}{ccccc} {}_p\mathbf{X}'_n\mathbf{X}_p & = & {}_p\mathbf{X}'_n\mathbf{H}_n\mathbf{X}_p & + & {}_p\mathbf{X}'_n[\mathbf{I}-\mathbf{H}]_n\mathbf{X}_p\,. \\ \mathbf{T} & = & \mathbf{B} & + & \mathbf{W}. \end{array} \quad (2) $$

We use the notation $\mathbf{Y} = \mathbf{HX}$ to denote the matrix of group means where the *k*th mean is repeatedly represented n_j times, $j = 1,\ldots,k$. Similarly, $(\mathbf{I} - \mathbf{H})\mathbf{X}$ is the matrix of residuals from the group means.

2. Ratio and Fit Criteria and How They Are Related

The classical form of CVA may be developed in two ways, which we term the ratio form and the fit form. These are described below.

2.1 The Ratio Form of CVA

Firstly, we may ask what linear combination $\mathbf{Xz}$ of the variables maximizes the ratio of the between-group variation to the within-group variation. Thus we require the solution to the following problem:

$$\textit{Ratio form}:\ \lambda = \max_{\mathbf{z}} \left(\frac{\mathbf{z}'\mathbf{Bz}}{\mathbf{z}'\mathbf{Wz}} \right). \tag{3}$$

The solution is given by the biggest eigenvalue λ and associated eigenvector $\mathbf{z}$ of

$$\mathbf{Bz} = \lambda\mathbf{Wz}. \tag{4}$$

The ratio form (3) explicitly seeks a one-dimensional vector solution $\mathbf{z}$. The scaling of $\mathbf{z}$ in (4) is an arbitrary identification constraint, though usually chosen so that $\mathbf{z'Wz} = 1$. We have $rank(\mathbf{B}) = k - 1$, at most, and $rank(\mathbf{W}) = min(n - k, p)$. When $\mathbf{W}$ is not of full rank p, there are problems; these are discussed in Section 3. The widespread use of multidimensional solutions to (4) based on suboptimal eigenvalues can be justified by the second form of CVA, described in Section 2.2.

A variant of the ratio form follows from noting that the problem

$$\max_{\mathbf{z}} \left(\mathbf{z}'\mathbf{Bz} \right) \text{ subject to the constraint } \mathbf{z'Wz} = 1$$

also leads to (4). Now, λ represents a Lagrange multiplier, rather than a ratio, and therefore the constraint is substantive and not merely a convenient identification constraint. Indeed, Healy and Goldstein (1976), in the context of determining optimal scores, pointed out that if other substantive constraints are used then a completely different maximum is found (for a discussion see Gower 1998). For example, with the constraint $\mathbf{z'1} = 1$ a maximum of $1/(\mathbf{1}'\mathbf{B}^{-1}\mathbf{1})$ occurs at $\mathbf{z} = \mathbf{B}^{-1}\mathbf{1}/(\mathbf{1}'\mathbf{B}^{-1}\mathbf{1})$, provided $\mathbf{1}'\mathbf{B}^{-1}\mathbf{1} \neq 0$. Rather confusingly, note that we are at liberty to use any arbitrary identification constraint, including $\mathbf{z'1} = 1$, to scale the unique eigenvector solution to (4). However, general Lagrangian constraints are not consistent with solutions to the CVA problem, which is not to say that they may not be relevant in other contexts not pursued here.

The above discussion alerts us to what might be appropriate forms of generalization. None would dispute that any generalization must include the basic form as a special case but this does not imply that a property of the basic form must necessarily be a property of a generalization. Thus, the basic form with vectors scaled so that $\mathbf{z'1} = 1$ is valid but if the scaling is made a substantive constraint there is only one solution to (4) and that is not the solution of (3). In the following we shall meet other instances where care has to be taken with generalization.

2.2. The Fit Form of CVA

In a second approach, we seek to fit a reduced rank matrix of rank r to linear combinations $\mathbf{YM}$ of the group means matrix $\mathbf{Y} = \mathbf{HX}$ (see (1) and following for notation). $\mathbf{M}$ is chosen so that the inner-product $\mathbf{Y} - \mathbf{M}'\mathbf{Y}'$ defines (i. e. generates) Mahalanobis distances between the group-means, requiring that $\mathbf{MM}' = \mathbf{W}^{-1}$. Transformed variables $\mathbf{XM}$ are known as canonical variables and their means $\mathbf{YM}$ as canonical means. Thus, in this form of CVA we must solve:

$$\textit{Fit form}: \quad \min_{\hat{\mathbf{Y}}} \| (\mathbf{Y} - \hat{\mathbf{Y}})\mathbf{M} \|^2 \text{ where } {}_p\mathbf{M}_p\mathbf{M}'_p = \mathbf{W}^{-1} \text{ and } rank(\hat{\mathbf{Y}}) = r. \quad (5)$$

As with the ratio form of CVA, it is initially assumed that $\mathbf{W}$ is of full rank, so having a proper inverse. Because the criterion (5) may be written as:

$$\min_{\hat{\mathbf{Y}}} trace(\mathbf{Y} - \hat{\mathbf{Y}})\mathbf{W}^{-1}(\mathbf{Y} - \hat{\mathbf{Y}})',$$

it is often referred to as a least-squares problem in the metric $\mathbf{W}$, giving a rank-r fit $\hat{\mathbf{Y}}$ to $\mathbf{Y}$ depending only on $\mathbf{W}$ and not the precise specification of $\mathbf{M}$. In the current context, we are more interested in the ordinary least-squares fit to the canonical means $\mathbf{YM}$ rather than to a weighted estimate of $\mathbf{Y}$. The solution to (5) is given by the ordinary Eckart-Young theorem (1936) as:

$$\hat{\mathbf{Y}}\mathbf{M} = \mathbf{U\Sigma J}_r\mathbf{V}' \text{ where } \mathbf{YM} = \mathbf{U\Sigma V}' \text{ is the SVD of } \mathbf{YM}, \quad (6)$$

and $\mathbf{J}_r$ is zero except for units in the r leading diagonal positions. The solution (6) is quite general and does not depend on the precise settings of $\mathbf{Y}$ and $\mathbf{M}$ whose particular forms are chosen here to bring the problem into the orbit of CVA. Thus, the procedure is (i) define a metric ${}_p\mathbf{M}_p\mathbf{M}'_p = \mathbf{W}^{-1}$ and then (ii) obtain an r-dimensional principal components fit (6) to the canonical means **YM**. We regard the above two-step procedure based on (6) as embodying the fundamental essence of CVA and endeavour to preserve it in our generalizations. Section 2.3 shows that the two step procedure is equivalent to the more familiar two-sided eigenvalue formulation, which remains a convenient vehicle for computation but nevertheless is secondary. We shall see that the two-step procedure is conceptually rich as a basis for developing generalizations that do not require **W** to be of full rank.

The two step procedure is all that is needed to describe the fit form of CVA but visualization plays an important part in presenting results and in interpretation. The basis of visualization is that the rows of $\hat{\mathbf{Y}}\mathbf{M}$ give the coordinates of an r-dimensional approximation embedded in a p-dimensional space. To get the simpler r-dimensional coordinates used in visualizations requires a rotation to principal axes, thus leaving inter-row distances unchanged, which then give rotated canonical variables **YMV** with their approximation $\hat{\mathbf{Y}}\mathbf{MV}$. Writing $\mathbf{Z} = \mathbf{MV}$ we have from (6):

$$\hat{\mathbf{Y}}\mathbf{Z} = \mathbf{YZJ}_r. \tag{7}$$

Thus (7) gives the required fit to the canonical means **YM**; the fit to the group means themselves **Y** are:

$$\hat{\mathbf{Y}} = \mathbf{YZJ}_r\mathbf{Z}^{-1}.$$

The canonical variables are **YZ** and we plot its first r columns $\mathbf{YZJ}_r$ as coordinates of the group-means. From (7) note that $\mathbf{YZJ}_r = \hat{\mathbf{Y}}\mathbf{Z} = \hat{\mathbf{Y}}\mathbf{ZJ}_r$ so that plots given by the original data **Y** and the fitted values $\hat{\mathbf{Y}}$ are the same. As well as plotting $\mathbf{YZJ}_r$ to give positions of the canonical means, we may also plot $\mathbf{XZJ}_r$ for the individual samples.

2.3 The Relationship Between the Ratio and Fit Forms

The algebra of the preceding section may be re-expressed. Starting from the SVD (6) we may proceed progressively to give:

$$\begin{aligned}
\mathbf{M}'(\mathbf{Y}'\mathbf{Y})\,\mathbf{M} &= \mathbf{V\Sigma}^2\mathbf{V}' \\
\mathbf{W}^{-1}(\mathbf{Y}'\mathbf{Y})\,\mathbf{MV} &= \mathbf{MV\Sigma}^2 \\
\mathbf{BZ} &= \mathbf{WZ\,\Sigma}^2,
\end{aligned} \tag{8}$$

an eigenvalue problem which, to be consistent with (5), requires that $\mathbf{M}'\mathbf{WM} = \mathbf{I}$ implying that that the eigenvectors **Z** of (8) should be normalized so that $\mathbf{Z}'\mathbf{WZ} = \mathbf{I}$. That (8) is in the form of the two-sided eigenvalue problem (4) establishes the link between the Ratio and Fit forms of CVA. In (6), **M** seems to be needed explicitly but (8) shows that knowledge of **W** suffices.

Computationally, it may be more convenient to solve the single eigenvalue problem (8) rather than the conceptually more transparent use of two steps, first defining the canonical variables, followed by a PCA (6). More importantly, (8) shows that the minimization of the matrix fitting criterion (5), in its one-dimensional form, is equivalent to that of maximizing (3), the ratio of quadratic forms

$$\frac{\mathbf{z}'\mathbf{B}\mathbf{z}}{\mathbf{z}'\mathbf{W}\mathbf{z}}.$$

Thus, the r-dimensional fit form of CVA includes the one-dimensional ratio form (see Fisher's Linear Discriminant Function), so justifying the extension to using multidimensional solutions for the ratio case. We opine that the ratio form is really only acceptable readily in its one-dimensional form. To justify multidimensional generalizations one has to appeal directly to the fit form, with its overt dependence on the Mahalanobis, or conceptually other, metric.

2.4 Miscellaneous Remarks on ANOVA and Weighting

We conclude Section 2 with some miscellaneous remarks on ANOVA in CVA and adaptations for operating with weighted or unweighted means.

The orthogonal breakdown:

$$\mathbf{YZ} = \hat{\mathbf{Y}}\mathbf{Z} + (\mathbf{Y} - \hat{\mathbf{Y}})\,\mathbf{Z} = \mathbf{YZJ}_r + \mathbf{YZ}(\mathbf{I} - \mathbf{J}_r)$$

gives inner products

$$\mathbf{YW}^{-1}\mathbf{Y}' = \hat{\mathbf{Y}}\mathbf{W}^{-1}\hat{\mathbf{Y}}' + (\mathbf{Y} - \hat{\mathbf{Y}})\mathbf{W}^{-1}(\mathbf{Y} - \hat{\mathbf{Y}})',$$

generating an orthogonal breakdown in terms of Mahalanobis squared-distances. This ANOVA, based on the PCA of the canonical variables, gives the fitted and residual components of the squared Mahalanobis distances between the means and is additional to the between/within ANOVA (2) of the untransformed data. Lubbe-Gardner, Le Roux and Gower (2008) discussed how it may be used to assess the quality of CVA fits.

We have defined (see (1) and following for notation) $\mathbf{Y}$ as ${}_n\mathbf{G}_k\mathbf{N}_k^{-1}\mathbf{G}'_n\mathbf{X}_p$ thus replicating the point for the kth canonical mean n_k times; this may be avoided by replacing $\mathbf{Y}$ by $\mathbf{Y}_1 = {}_k\mathbf{N}_k^{-1}\mathbf{G}'_n\mathbf{X}_p$. It may be argued that it is better to plot the canonical means relative to the principle axes of the canonical means unweighted by their sample sizes. This requires the eigenvectors of

$$\mathbf{M}'\mathbf{Y}_1'\mathbf{Y}_1\mathbf{M} = \mathbf{M}'\,{}_p\mathbf{X}_n'\mathbf{G}_k\mathbf{N}_k^{-2}\mathbf{G}_n'\mathbf{X}_p\mathbf{M},$$

rather than of

$$\mathbf{M}'\mathbf{Y}'\mathbf{Y}\mathbf{M} = \mathbf{M}_p'\mathbf{X}_n'\mathbf{G}_k\mathbf{N}_k^{-1}\mathbf{G}_n'\mathbf{X}_p\mathbf{M} = \mathbf{M}_p'\mathbf{B}_p\mathbf{M} = \mathbf{\Sigma}^2,$$

and is automatically provided by replacing $\mathbf{Y}$ by $\mathbf{Y}_1$ in (6).

3. W Not of Full Rank

So far, we have assumed that $\mathbf{W}$ is of full rank. This condition is usually satisfied in the classical case, where the number of samples n is greater than the number of variables p. Of increasing importance is the case where $n << p$, in which case ${}_p\mathbf{W}_p$ is singular so we cannot have $\mathbf{MM}' = \mathbf{W}^{-1}$ as before, a key condition underpinning the link with Mahalanobis distance. To make progress with this new situation we note that for any nonsingular $\mathbf{L}$, (3) and (5) may be rewritten:

$$\max_{\mathbf{z}}\left(\frac{\mathbf{z}'\mathbf{Bz}}{\mathbf{z}'\mathbf{Wz}}\right) = \max_{\mathbf{L}^{-1}\mathbf{z}}\left(\frac{\mathbf{z}'\mathbf{L}'^{-1}(\mathbf{L}'\mathbf{BL})\mathbf{L}^{-1}\mathbf{z}}{\mathbf{z}'\mathbf{L}'^{-1}(\mathbf{L}'\mathbf{WL})\mathbf{L}^{-1}\mathbf{z}}\right)$$

and

$$\min_{\hat{\mathbf{Y}}} \| (\mathbf{Y}-\hat{\mathbf{Y}})\mathbf{M} \|^2 = \min_{\hat{\mathbf{Y}}\mathbf{L}} \| (\mathbf{YL}-\hat{\mathbf{Y}}\mathbf{L})\mathbf{L}^{-1}\mathbf{M} \|^2 .$$

We may seek a nonsingular transformation $\mathbf{L}$ that simplifies both forms (3) and (5) of the CVA problem. Such a transformation exists, as is discussed below. The non-singularity condition not only ensures the existence of $\mathbf{L}^{-1}$ but also allows the back-transformation of $\mathbf{L}^{-1}\mathbf{z}$ or $\mathbf{YL}$, as appropriate. The precise choice of $\mathbf{L}$ makes little difference to the ratio form of the CVA criterion but it is crucial to the fit form, where, as we shall see, it is fundamental in defining the metric. In analogy with the full rank case, it seems natural to require that $\mathbf{LL}'$ be a generalized inverse of $\mathbf{W}$. This is satisfied when $\mathbf{L}'\mathbf{WL} = \mathbf{J}$, a diagonal matrix of zeros and units, but there are many g-inverses each defining different metrics (see Appendix A). The remainder of this section is concerned with an examination of how major variants of canonical analysis arise from metrics based on different g-inverses of $\mathbf{W}$. Two metrics (the GCF-metric and the MP-metric explained in the following and defined in Appendix A) are examined in some detail but there are further possibilities.

In analogy with the basic canonical form of the full rank case, we appeal to the general diagonal canonical forms (GCF) of $\mathbf{B}$ and $\mathbf{W}$, discussed by Albers, Critchley, and Gower (2011). The GCF finds a nonsingular transformation $\mathbf{L}$ that simultaneously reduces $\mathbf{B}$ and $\mathbf{W}$ to diagonal form and, as a consequence of the summation in (2), also diagonalizes $\mathbf{T}$.

This choice of **L** gives a useful simplification. The full generality of the GCF allows **B** to be indefinite, but here we need only a simpler form, in which **B** and **W** are p.s.d. (including the classical case where either or both are definite). Details of explicit expressions for the transformations are given in Appendix A. Fortunately, knowledge that the transformation exists suffices for establishing the main results, although the detailed form of the transformation is needed for computational purposes. Either **B** or **W** can be transformed to diagonal form with unit/zero diagonal values. We transform **W** to unit/zero form, written **J**, and **B** to diagonal form, written $\boldsymbol{\Gamma}$. It follows therefore (Appendix A) that ${}_p\mathbf{L}_p\mathbf{L}'_p = {}_p\mathbf{W}^-_p$ is a g-inverse of **W** analogous to the ordinary inverse ${}_p\mathbf{L}_p\mathbf{L}'_p = {}_p\mathbf{W}^{-1}_p$ of the full-rank case. Although the appeal to the GCF gives a useful simplified reparameterization of the optimization problems, it does not add any new constraints. In particular, the g-inverse result arises as a consequence of the transformation and is not an imposed constraint. Fundamentally, **MM'** remains an arbitrary metric but to retain the connection with Mahalanobis distance it should be related to the inverse of **W** and when **W** is singular it is natural to replace $\mathbf{W}^{-1}$ by a g-inverse $\mathbf{W}^-$. One possibility is to identify **M** with **L**. This gives the GCF-metric, the first of the two metrics that we examine in detail. Thus, we may replace (3) and (5) by:

$$\textit{Ratio form}:\ \max_{\boldsymbol{\zeta}}\left(\frac{\boldsymbol{\zeta}'\boldsymbol{\Gamma}\boldsymbol{\zeta}}{\boldsymbol{\zeta}'\boldsymbol{\zeta}}\right)\ \text{where}\ \boldsymbol{\zeta}=\mathbf{L}^{-1}\mathbf{z} \tag{9}$$

and

$$\textit{Fit form}:\ \min_{\hat{\mathbf{Y}}}\left\|(\mathbf{Y}-\hat{\mathbf{Y}})\mathbf{L}\right\|^2, \tag{10}$$

where $rank(\hat{\mathbf{Y}}) = r$ and ${}_p\mathbf{L}_p\mathbf{L}'_p = {}_p\mathbf{W}^-_p$; **L** is given (Appendix A) and is nonsingular.

The relationships between the ranks of the matrices involved underpin structural differences that are fundamental to the understanding of our subsequent development. In general

$$\text{rank}(\mathbf{T}) = \min(n-1, p),\ \text{rank}(\mathbf{W}) = \min(n-k, p)\ \text{and}\ \text{rank}(\mathbf{B}) = k-1. \tag{11}$$

Collinearities may reduce these ranks, but they may be adjusted accordingly without substantive effect.

The remainder of this section analyzes three mutually exclusive possibilities: (i) $p \le n-k$ (Section 3.1) (ii) $n-k < p < n-1$ (Section 3.3) and (iii) $n-1 \le p$ (Section 3.2). Throughout, we use the notation $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(0)}$ to denote matrices in the range and null spaces, respectively, of **W** so freeing the positions of suffices for showing matrix dimensions.

3.1 The GCF When Rank (W) = p (The Mahalanobis Metric)

In the classical case where $p \le n - k$ then rank($\mathbf{W}$) = rank($\mathbf{T}$) = p and $\mathbf{T}$ and $\mathbf{W}$ are both of full rank and so have no null space. In this case $\mathbf{W}$ has a unique inverse and the possibility of using different metrics based on different g-inverses does not arise; the Mahalanobis and GCF-metric coincide. We assume that $rank(\mathbf{B}) = k - 1$, thus avoiding tedious, but essentially trivial, caveats to cover the possibility of collinearities among the group-means. Then, $\mathbf{W}$ has no null space and the matrix $\mathbf{U}^{(0)}$ of Appendix A vanishes so that $\mathbf{\Gamma}^{(1)}$ is of necessity in the range space of $\mathbf{W}$ and the GCF simplifies to the classical form, disregarding nondiagonal zero blocks.

$$\begin{array}{ccccc} \begin{pmatrix} \mathbf{\Gamma}^{(1)}_{k-1} + \mathbf{I}_{k-1} & \\ & \mathbf{I}_{p-k+1} \end{pmatrix} & = & \begin{pmatrix} \mathbf{\Gamma}^{(1)}_{k-1} & \\ & \mathbf{0} \end{pmatrix} & + & \begin{pmatrix} \mathbf{I}_{k-1} & \\ & \mathbf{I}_{p-k+1} \end{pmatrix} \\ \mathbf{L'TL} & = & \mathbf{L'BL} & + & \mathbf{L'WL} \\ \mathbf{L'X'XL} & = & \mathbf{L'X'HXL} & + & \mathbf{L'X'(I-H)XL} \end{array} \quad (12)$$

The canonical form (12) shows that maximizing $\mathbf{L'BL}$ in r dimensions is achieved by choosing the solution associated with the r largest values γ_i of $\mathbf{\Gamma}^{(1)}$. The ratio and best-fitting matrix forms coincide for $r = 1$, while the best-fitting matrix form is valid for $r > 1$, as we saw in Section 2. As required by (5),

$$\mathbf{L'WL} = \mathbf{I} \text{ and } \mathbf{L'BL} = \begin{pmatrix} \mathbf{\Gamma}^{(1)} & \\ & \mathbf{0} \end{pmatrix},$$

from which

$$(\mathbf{L'BL}) = (\mathbf{L'WL}) \begin{pmatrix} \mathbf{\Gamma}^{(1)} & \\ & \mathbf{0} \end{pmatrix},$$

showing, because $\mathbf{L}$ is nonsingular, that $\mathbf{L}$ may be found from the two-sided eigenvalue problem

$$\mathbf{BL} = \mathbf{WL} \begin{pmatrix} \mathbf{\Gamma}^{(1)} & \\ & \mathbf{0} \end{pmatrix},$$

as with (4) and the special form taken by the general solution for $\mathbf{L}$ given in Appendix A. Because $\mathbf{L'BL}$ is diagonal, the canonical variables are referred to their principal axes. The link with Mahalanobis distance follows from the between group inner product $\mathbf{HX(LL')X'H} = \mathbf{HX(W^{-1})X'H}$. In this case, the GCF reproduces the classical results.

3.2 The GCF When Rank (W) = n – k<<p (The GCF-Metric)

When $n - 1 \leq p$ then rank(**T**) = $n - 1$, rank(**W**) = $n - k$ and the GCF says that there is a nonsingular transformation **XL** of **X** that gives the decomposition:

$$\begin{pmatrix} \mathbf{I}_{n-k} & & \\ & \mathbf{\Gamma}^{(0)}_{k-1} & \\ & & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & & \\ & \mathbf{\Gamma}^{(0)}_{k-1} & \\ & & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_{n-k} & & \\ & \mathbf{0} & \\ & & \mathbf{0} \end{pmatrix} \tag{13}$$

$$\begin{aligned} &\mathbf{L'T\,L} &&= \mathbf{L'B\,L} &&+\ \mathbf{L'WL} \\ &\mathbf{L'X'X\,L} &&= \mathbf{L'X'HX\,L} &&+\ \mathbf{L'X'(I-H)X\,L} \end{aligned}$$

The decomposition (13) differs from that of (12) and needs some explanation. Firstly, if **u** is a null vector of **T** then **X'Xu** = 0, which implies that **Xu** = 0 and hence that **Bu** = **Wu** = 0. This explains the final row of (13): all null vectors of **T** are also null vectors of **B** and **W**; we term this their common null-space. The common null space is irrelevant to canonical analysis, so is ignored in the following. From (13), both **B** and **W** have additional null vectors that are discussed in the following. Because rank(**T**) = $n - 1$ and rank(**B**) = $k - 1$ and rank(**W**) = $n - k$ then the diagonal matrix $\mathbf{\Gamma}^{(0)}_{k-1}$ must lie entirely in what we term the intersection space, strictly the intersection of the range space of **T** with the null space of **W**. This result implies that, in the reduce rank form of CVA, the diagonal matrix $\mathbf{\Gamma}^{(1)}$ of the GCF (Appendix A) vanishes, a property that is at the root of the difficulties that we are about to describe. Thus, the between and within components are in orthogonal subspaces of the total variation, implying that maximizing the ratio **z'Bz/z'Wz** is now meaningless because $\zeta'\Gamma\zeta$ *and* $\zeta'\zeta$ of (9) are independent, so we may maximize **z'Bz** independently of **z'Wz** and we may minimize **z'Wz** independently of **z'Bz**. Indeed, we may make either term zero by selecting ζ to lie in one or the other null space. Albers, Critchley and Gower (2011) discuss further ramifications of this problem. It is important to recognize that this result is a fundamental property of the ratio criterion and does not depend on **L** being derived from the GCF. Similar difficulties stem from using criteria based on *trace*($\mathbf{T}^{-}\mathbf{B}$) or similar. Thus, the Ratio form of CVA runs into difficulties when $p >> n$ and therefore we do not pursue it further. However, the fit form (5) continues to have meaning. Indeed it may have more than one meaning depending on whether a solution is sought within the intersection space or in the range space of **W**, as is shown in the following and discussed further in Section 4. Thus, we follow the two-stage procedure discussed in Section 2.2 of (i) establishing a metric, followed by (ii) a PCA of the means of the canonical variables. The differences in fit are related to choices of metric, two of which are explored in some detail: the GCF metric (Section 3.2.1) and the Moore-Penrose metric, MP (Section 3.2.2).

3.2.1 Fit Form in the Intersection Space of W (The GCF-Metric)

Continuing for the present with the general canonical form (13), the solution to the generalized fit-form of CVA (11) is given, as before, by the Eckart-Young theorem applied to (6) as:

$$\hat{\mathbf{Y}}\mathbf{L} = \mathbf{U\Sigma J}_r\mathbf{V}' \text{ where } \mathbf{YL} = \mathbf{U\Sigma V}' \text{ is the SVD of } \mathbf{YL}, \tag{14}$$

where now $\mathbf{L}$ is given by (A2) of the appendix. Recalling the superscript notation denoting range and null spaces of $\mathbf{W}$, the leading term of $\mathbf{L'BL}$ in (13) shows that $\mathbf{HXL}^{(1)} = \mathbf{0}$, so $\mathbf{YL} = (\mathbf{0}, \mathbf{HXL}^{(0)})$; the canonical means are entirely in the intersection space. From the null space of $\mathbf{W}$ we also have that $(\mathbf{I} - \mathbf{H})\mathbf{XL}^{(0)} = \mathbf{0}$ so we may simplify the expression for the canonical means to give $\mathbf{YL} = (\mathbf{0}, \mathbf{XL}^{(0)})$.

From (14),

$$\hat{\mathbf{Y}}\mathbf{LV} = \mathbf{U\Sigma J}_r$$
$$\hat{\mathbf{Y}}\mathbf{Z} = \mathbf{YZJ}_r\ ,$$

giving the fitted values in r dimensions as:

$$\hat{\mathbf{Y}} = \mathbf{YZJ}_r\mathbf{Z}^{-1},$$

where $\mathbf{Z} = \mathbf{LV}$, with valid inverse because $\mathbf{L}$ is nonsingular. Equation (13) shows that now $\mathbf{L'BL} = (\mathbf{I}_{n-1} - \mathbf{I}_{n-k})\ \boldsymbol{\Gamma}^{(0)} = (\mathbf{I} - \mathbf{L'WL})\,\boldsymbol{\Gamma}^{(0)}$ which does not simplify to give the classical two-sided eigenvalue solution (Section 3.1), now requiring that $\mathbf{L}$, and hence $\mathbf{Z}$, are calculated from the GCF as given explicitly in Appendix A. The main differences from the full rank case are:

(a) The two-sided eigenvalue formulation is replaced by the GCF.

(b) Variation between the canonical means is orthogonal to variation within groups and it follows that only between-group distances in the intersection space can be defined.

(c) The inverse metric $\mathbf{LL}' = \mathbf{W}^{-1}$ is replaced by a g-inverse $\mathbf{LL}' = \mathbf{W}^{-}$ (Appendix A).

The canonical variables are $\mathbf{YZ}$ and we plot the first r columns $\mathbf{YZJ}_r$ as coordinates. Note that $\mathbf{YZJ}_r = \hat{\mathbf{Y}}\mathbf{Z} = \hat{\mathbf{Y}}\mathbf{ZJ}_r$ so the original data $\mathbf{Y}$ and the fitted values $\hat{\mathbf{Y}}$ generate the same plots. Now, we have the orthogonal breakdown:

$$\mathbf{YZ} = \hat{\mathbf{Y}}\mathbf{Z} + (\mathbf{Y} - \hat{\mathbf{Y}})\ \mathbf{Z} = \mathbf{YZJ}_r + \mathbf{YZ}(\mathbf{I} - \mathbf{J}_r)\ ,$$

from which we derive

$$\mathbf{YW}^{-}\mathbf{Y}' = \hat{\mathbf{Y}}\mathbf{W}^{-}\hat{\mathbf{Y}}' + (\mathbf{Y} - \hat{\mathbf{Y}})\mathbf{W}^{-}(\mathbf{Y} - \hat{\mathbf{Y}})'\ ,$$

the inner products giving a breakdown in Mahalanobis distances, generalized for singular $\mathbf{W}$, that is similar to the full-rank case (Section 2.4). Thus, the ordinary Euclidean distances between the canonical means are generalized Mahalanobis distances.

We have seen that the distances between the canonical group-means derive from their coordinates $\mathbf{N}^{-1}\mathbf{G}'\mathbf{XL}^{(0)}$ in the intersection space. These distances are readily computed, especially with the help of results given in Appendix B. However, it is informative to have an algebraic expression for the derived distances. After extensive algebraic manipulations Gower and Albers (2011) showed that when $k = 2$ the distance d_{12} is given by:

$$d_{12}^{2} = \frac{n^{2}}{n_1^2 n_2^2}[(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\boldsymbol{\Lambda}^{-2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)]^{-1}\ .$$

and, more generally, provided $rank(\mathbf{HX}) = k - 1$, the distance d_{ij} between the means of the ith and jth group means is given by:

$$d_{ij}^{2} = \frac{(n_i + n_j)^2}{n_i^2 n_j^2}\left[\left(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\right)'\boldsymbol{\Lambda}^{-1}(\mathbf{I} - \mathbf{R}_{ij})\boldsymbol{\Lambda}^{-1}\left(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\right)\right]^{-1}$$

where $\mathbf{R}_{ij}$ represents orthogonal projection onto the space spanned by the remaining $(k - 2)$ means. In these expressions $\boldsymbol{\Lambda}$ represents the non-zero eigenvalues of $\mathbf{T}$. Although these results recall Mahalanobis distance, albeit inverted, it should be stressed that the means in the intersection space lie in the null space of $\mathbf{W}$. Consequently, there is no within-group variation and Mahalanobis distance is undefined and the usual probabilistic basis for discrimination is not available. Further, unlike Mahalanobis distance, the distances d_{ij} are not invariant to the scaling of $\mathbf{X}$, so require preliminary normalization, as in PCA.

Because of the complicated nature of $\mathbf{L}^{(1)}$ in (A2) involving $\mathbf{U}^{(0)}$, the within-group canonical coordinates $(\mathbf{I} - \mathbf{H})\mathbf{XL}$ generate inner products that are not invariant to the choice of $\mathbf{U}^{(0)}$ and neither are the associated within-group distances (see remarks after (A7)). Indeed, the use of $\mathbf{U}^{(0)}$ in the GCF is arbitrary to the extent that it may be replaced by $\mathbf{U}^{(0)}\mathbf{E}$ for any non-singular matrix $\mathbf{E}$, thus affecting between group distances. Now, $\mathbf{U}'^{(0)}\mathbf{BU}^{(0)}$ $= \boldsymbol{\Gamma}$ becomes $\mathbf{E}'\boldsymbol{\Gamma}\mathbf{E}$ which is no longer diagonal. However, it can easily be diagonalized by a PCA or we could even choose $\mathbf{E} = \boldsymbol{\Gamma}_{k-1}^{(0)-\frac{1}{2}}$ in which case we have diagonalized to a unit matrix. This modified form of **L** generates a metric giving equal distances between groups, so the group means are, unhelpfully, at the vertices of a (k – 1)-dimensional regular simplex, while **LL'** remains a g-inverse of **W** satisfying **L'WL = J**. Thus, the possibilities to be derived from full unit diagonalization are unhelpful. We conclude that while the canonical variables in the intersection space are worth consideration, the full canonical analysis derived from the GCF needs the further consideration discussed in Section 3.2.2. For these reasons, we consider only **E** = **I**, which does not suffer from these disadvantages and gives a between group distance that depends only on $\mathbf{U}^{(1)}$.

The choice of transformation that uses the orthonormal matrix $\mathbf{U}^{(0)}$ is termed Null Linear Discriminant Analysis (NLDA) in the Pattern Analysis and Machine Learning literature, because it is based on transformations of **X** and **HX** in the null space of **(I –H)X** that happen to be orthonormal (see e.g. Ye and Xiong 2006; Ye 2005; Chen, Liao, Ko, Lin, and Yu 2000). This may be contrasted with Orthogonal LDA (OLDA) which seeks orthogonally constrained transformations of **X**. Not surprisingly, the two approaches are linked. This is easily seen, by recalling that $(\mathbf{I} - \mathbf{H})\mathbf{X}\mathbf{L}^{(0)} = \mathbf{0}$ so that an NLDA of $\mathbf{HX}^{(0)}\mathbf{L}^{(0)}$ is bound to give a solution to an OLDA of $\mathbf{XL}^{(0)}$ and so is a candidate for an optimal solution to OLDA. In the classical case, where there is no null space, no NLDA candidate solution is available; nevertheless OLDA will have an optimal solution. It follows that in the classical case, OLDA must have an optimal solution in the range space of **W** (equivalently **T**) and that this solution satisfies the imposed substantive orthonormal constraint. Thus, OLDA adds to the examples discussed in Section 2.1 where a substantive constraint conflicts with the basic CVA criterion. As mentioned in Section 2.1, such a conflict does not necessarily imply that substantively constrained criteria do not have applications beyond CVA.

3.2.1.1 Some Identities and Between-Group Distance in the Intersection Space. The canonical form (13) gives some identities of interest. Writing $\mathbf{L} = \left({}_{\mathbf{p}}\mathbf{L}^{(1)}_{\mathbf{n-k}}, {}_{\mathbf{p}}\mathbf{L}^{(0)}_{\mathbf{p-n+k}}\right)$, partitioned conformably with (13), then $\mathbf{XL}^{(1)}$ generates the coordinates of the canonical variables within-groups and $\mathbf{XL}^{(0)}$ generates the coordinates of the canonical variables between-groups. In the latter case the coordinates of the kth group-mean are repeated n_k times. From the individual terms of (13) we have that:

$$\mathbf{L}'\mathbf{WL} = \begin{pmatrix} \mathbf{L}'^{(1)}\mathbf{WL}^{(1)} & \\ & \mathbf{L}'^{(0)}\mathbf{WL}^{(0)} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{\mathbf{n-k}} & \\ & \mathbf{0} \end{pmatrix}$$

$$\mathbf{L'BL} = \begin{pmatrix} \mathbf{L'^{(1)}BL^{(1)}} & \\ & \mathbf{L'^{(0)}BL^{(0)}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \\ & \mathbf{\Gamma}_{k-1}^{(0)} \end{pmatrix}$$

$$\mathbf{L'TL} = \begin{pmatrix} \mathbf{L'^{(1)}X'XL^{(1)}} & \\ & \mathbf{L'^{(0)}X'XL^{(0)}} \end{pmatrix} = \begin{pmatrix} \mathbf{I_{n-k}} & \\ & \mathbf{\Gamma}_{k-1}^{(0)} \end{pmatrix}$$

giving the following identities:

$$\mathbf{L'^{(1)}X'XL^{(1)}} = \mathbf{I_{n-k}}$$

$$\mathbf{L'^{(1)}X'HXL^{(1)}} = \mathbf{0} \Rightarrow \mathbf{HXL^{(1)}} = \mathbf{0} \Rightarrow \mathbf{G'XL^{(1)}} = \mathbf{0}$$

$$\mathbf{L'^{(0)}X'XL^{(0)}} = \mathbf{L'^{(0)}X'HXL^{(0)}} = \mathbf{\Gamma}_{k-1}^{(0)}$$

$$\mathbf{L'^{(0)}X'(I-H)XL^{(0)}} = \mathbf{0} \Rightarrow \mathbf{(I-H)XL^{(0)}} = \mathbf{0}\ .$$

3.2.2 Fit Form in the Range Space of W (The MP-Metric)

In the classical case of 3.1 the intersection space is null so the full-rank solution is necessarily in the range space of **W**. By contrast, the solution in the intersection space given in Section 3.2.1 using the GCF-metric is orthogonal to the range space and so is of a different nature. When **W** is not of full rank, an interesting solution (the MP-metric, see Appendix A2) exists in the range space, that does not depend on the full GCF. In this section we define **L** by (A4), $\mathbf{L}^{(1)} = \mathbf{U}^{(1)}\mathbf{\Delta}^{-1}, \mathbf{L}^{(0)} = \mathbf{U}^{(0)}$, which diagonalizes **W** to **J** and hence **LL**' remains a g-inverse of **W**. It gives the following canonical form where **W** has unit/zero diagonal.

$$\begin{pmatrix} \mathbf{C}_{11}+\mathbf{J} & \mathbf{C}_{10} \\ \mathbf{C}_{01} & \mathbf{C}_{00} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{10} \\ \mathbf{C}_{01} & \mathbf{C}_{00} \end{pmatrix} + \begin{pmatrix} \mathbf{J} & \\ & \mathbf{0} \end{pmatrix}$$

$$\begin{array}{lll} \mathbf{L'TL} & = \mathbf{L'BL} & + \mathbf{L'WL} \\ \mathbf{L'X'XL} & = \mathbf{L'X'HXL} & + \mathbf{L'X'(I-H)XL}\ . \end{array}$$

This transformation generates the between-group matrix

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{10} \\ \mathbf{C}_{01} & \mathbf{C}_{00} \end{pmatrix}$$

of Appendix A. If we went on to use orthogonal transformations to diagonalize **C**, we would arrive at the GCF form of Section 3.2.1 but here we consider the form where **L′BL** = **C**, which is not diagonal, while **L′WL** generates a generalized Mahalanobis metric. From Appendix A, $\mathbf{C}_{11} = \mathbf{\Delta}^{-1}\mathbf{U}'^{(1)}\mathbf{B}\mathbf{U}^{(1)}\mathbf{\Delta}^{-1}$ is the part of **C** that is entirely in the range space of **W**. Thus, $\mathbf{C}_{11}$ now derives from canonical means $\mathbf{HXU}^{(1)}\mathbf{\Delta}^{-1}$ and it is this matrix that generates an r-dimensional fit, obtained in the usual way from its SVD or PCA. The Mahalonobis distances derived from $\mathbf{C}_{11}$ do not depend overtly on $\mathbf{U}^{(0)}$, but see the following paragraph below. Moreover, the within-group coordinates are $(\mathbf{I} - \mathbf{H})\mathbf{XL}^{(1)}$, so the within-group distances also are invariant to the choice of g-inverse (See remarks following (A7)). This canonical analysis is very close to the classical analysis of defining a metric $\mathbf{L}^{(1)}\mathbf{L}'^{(1)}$ followed by a PCA of $\mathbf{HXL}^{(1)}$ and is often used (see e.g. Krzanowski, Jonathan, McCarthy, and Thomas 1995; Mardia 1977; and Rao and Yanai 1979).

We note that diagonalization of $\mathbf{C}_{00}$ in the GCF entailed the automatic elimination of $\mathbf{C}_{01}$, $\mathbf{C}_{10}$ and $\mathbf{C}_{11}$ in (13). The diagonalization of $\mathbf{C}_{11}$ does not entail the elimination of $\mathbf{C}_{01}$, $\mathbf{C}_{10}$ and $\mathbf{C}_{00}$ so any information contained in these matrices is ignored. However, the result $\mathbf{C}_{11} - \mathbf{C}_{10}\mathbf{C}_{00}^{-1}\mathbf{C}_{01} = \mathbf{0}$ remains valid, showing that the intersection space continues to play a major part in the analysis of $\mathbf{C}_{11}$. It also shows that only the eigendecomposition of the $(k-1)\times(k-1)$ matrix:

$$\left(\mathbf{\Gamma}^{(0)}\right)^{-1/2}\mathbf{Q}^{(1)}\mathbf{C}_{01}\mathbf{C}_{10}\mathbf{Q}'^{(1)}\left(\mathbf{\Gamma}^{(0)}\right)^{-1/2}$$

is required, as described in Appendix B.

3.2.2.1 Other Metrics Derived from the MP-Metric. Alternatively, we may derive the canonical means from $\mathbf{C}_{00}$ or from the whole of $\mathbf{C}$. Note that $\mathbf{C}$, $\mathbf{C}_{11}$ and $\mathbf{C}_{00}$ all have rank $k - 1$, and each of the rank r fits may be displayed relative to principal axes. The solution based on $\mathbf{C}_{11}$ was discussed in Section 3.2.2, while that based on $\mathbf{C}_{00}$ is precisely the same as in Section 3.2.1, so is the same as for the GCF-metric and is independent of $\mathbf{U}^{(0)}$, as found there. When derived from the whole of $\mathbf{C}$, the solution is in the range space of $\mathbf{T}$ and so contains the intersection space. Then, the individual terms of $\mathbf{C}$ given after (A1) may be rewritten as:

$$\begin{pmatrix}\Delta^{-1}\mathbf{U}'^{(1)}\\ \mathbf{U}'^{(0)}\end{pmatrix}\mathbf{X}'\mathbf{HX}\begin{pmatrix}\mathbf{U}^{(1)}\Delta^{-1} & \mathbf{U}^{(0)}\end{pmatrix} = \begin{pmatrix}\Delta^{-1}\mathbf{U}'^{(1)}\mathbf{X}'\mathbf{H}\\ \mathbf{U}'^{(0)}\mathbf{X}'\end{pmatrix}\begin{pmatrix}\mathbf{HXU}^{(1)}\Delta^{-1} & \mathbf{XU}^{(0)}\end{pmatrix},$$

where we have used the result (Section 3.2.1.1) that $\mathbf{HXU}^{(0)} = \mathbf{XU}^{(0)}$. The coordinates $(\mathbf{HXU}^{(1)}\Delta^{-1} \quad \mathbf{XU}^{(0)})$ generate the inner-product:

$$\begin{pmatrix}\mathbf{HXU}^{(1)}\Delta^{-1} & \mathbf{XU}^{(0)}\end{pmatrix}\begin{pmatrix}\Delta^{-1}\mathbf{U}'^{(1)}\mathbf{X}'\mathbf{H}\\ \mathbf{U}'^{(0)}\mathbf{X}'\end{pmatrix}$$
$$= \mathbf{HXU}^{(1)}\Delta^{-2}\mathbf{U}'^{(1)}\mathbf{X}'\mathbf{H} + \mathbf{XU}^{(0)}\mathbf{U}'^{(0)}\mathbf{X}',$$

the last term of which may be written $\mathbf{X}(\mathbf{I} - \mathbf{U}^{(1)}\mathbf{U}'^{(1)})\mathbf{X}'$. This establishes that the derived distances do not depend on $\mathbf{U}^{(0)}$ and are invariant to the choice of generalized inverse (A7). Because the canonical variables are $\mathbf{X}(\mathbf{L}^{(1)}, \mathbf{L}^{(0)})$ it follows that the distances c_{ij} generated by the analysis of $\mathbf{C}$ have the form $c_{ij}^2 = D_{ij}^2 + d_{ij}^2$ where D_{ij}^2 are the Mahalanobis distances given by the MP-metric analysis of $\mathbf{C}_{11}$ and d_{ij}^2 are distances given by the GCF-metric in the intersection space derived from the analysis of $\mathbf{C}_{00}$ and discussed in Section 3.2.1.

3.3 The GCF When Rank(W) = n – k < p and Rank(T) = p < n – 1

There remains the case where $n - k < p < n - 1$. To establish the fundamental structure of this case, we return to the GCF. Then, $rank(\mathbf{W}) = n - k < p$ and $rank(\mathbf{T}) = p < n - 1$ and we demonstrate that $\mathbf{\Gamma}$ is partly in the null space of $\mathbf{W}$ and partly not. This occurs in a window, usually small, of width $k - 1$, e.g. $p = 7$, $n = 9$, $k = 3$ or $p = 16$, $n = 20$, $k = 6$. We illustrate the latter case in (16), where the suffices denote rank:

$$\begin{pmatrix} \mathbf{I}_3 + \mathbf{\Gamma}_3^{(1)} & & \\ & \mathbf{I}_{11} & \\ & & \mathbf{\Gamma}_2^{(0)} \end{pmatrix} = \begin{pmatrix} \mathbf{\Gamma}_3^{(1)} & & \\ & \mathbf{0}_{11} & \\ & & \mathbf{\Gamma}_2^{(0)} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_3 & & \\ & \mathbf{I}_{11} & \\ & & \mathbf{0}_2 \end{pmatrix} \quad (16)$$
$$\mathbf{L'TL} \qquad = \qquad \mathbf{L'BL} \qquad + \qquad \mathbf{L'WL}$$

Because rank ($\mathbf{T}$) = 16 and rank ($\mathbf{W}$) = 14 there are only two dimensions in the null space of $\mathbf{W}$ available for the 5 dimensions of $\mathbf{\Gamma}$, which therefore has to be split into $\mathbf{\Gamma}_2^{(0)}$ in the null space and $\mathbf{\Gamma}_3^{(1)}$ in the range space of $\mathbf{W}$. We have not followed up this case in detail beyond noting that the ratio $\mathbf{z'Bz/z'Wz}$ has a maximum, max $\mathbf{\Gamma}^{(1)}$, in the range space shared by $\mathbf{B}$ and $\mathbf{W}$ (in the example three-dimensional). If, however, we include the intersection space, the maximum increases to max $\mathbf{\Gamma}^{(1)}$ + trace $\mathbf{\Gamma}^{(0)}$. This solution need not necessarily be included in the fit form of the CVA problem; it depends whether or not the biggest γ occurs in $\mathbf{\Gamma}_3^{(1)}$ or $\mathbf{\Gamma}_2^{(0)}$. In general, the best r-dimensional fit may lie fully in the range space of $\mathbf{W}$ or fully in the part of the null space of $\mathbf{W}$ that is in the range space of $\mathbf{T}$, or it may straddle both spaces.

4. Example

This example relates to four groups each with six samples. The data, based on previous studies by Queen, Wright, and Albers (2007), and Queen and Albers (2009), are vehicle flow counts on different sections of the M25/A296 motorway network in Kent, United Kingdom. The groups refer to the $k = 4$ slip roads between these motorways and the $n_1 = n_2 = n_3 = n_4 = 6$ samples refer to different weeks, each about a month apart. The variables are counts of vehicle movements taken in the $p = 168$ hours for each week.

Figure 1a shows the analysis, based on $\mathbf{C}_{11}$, in the range space with the usual within-group dispersion around the group-means. This type of analysis is close to classical CVA but confined to the dimensions that do not have null within-group dispersion. Clearly there is good separation between the canonical means with moderate within-group variation, apart from a single outlier (of the type 'Δ') arising from a known traffic incident.

After initial normalization of **X**, Figure 1b shows the two-dimensional fit to the four means in the three-dimensional intersection space; we would get an exact fit in three dimensions. Of necessity, there is no within-group variation in the null-space of **W.** The means are clearly differently disposed to those of Figure 1a, even allowing for rotational indeterminacy.

Figure 1c combines the previous two analyses. It shows the same means as in Figure 1b but with the within-group dispersion from Figure 1a. Thus, although a two-dimensional approximation, this is a four-dimensional representation. The two dimensions within-groups are orthogonal to the two between groups dimensions.

5. Discussion

Perhaps one of the most interesting findings has been the non-equivalence of ratio and fit criteria, even in the classical case. Especially interesting is the transition through the window of case (ii). The GCF has been useful for examining basic multidimensional structure, giving a complete analysis of the behaviour of the ratio criterion for variants of the CVA problem. It has been successful in isolating the role of suitable metrics for examining the fit versions of the criterion. The simpler choice of **L** given by (A4) and the MP-metric being among the more useful in giving Mahalanobis distances that are independent of the arbitrary nature of $\mathbf{U}^{(0)}$.

An interesting feature concerns the orthogonality of the group-mean space and the within groups space when $p >> n - k$. Then, the GCF shows that two sets of linear combinations can always be found, that entirely separate the between and within group variations; one set is entirely in the range space of **W** and the other entirely in the intersection space. The groups can always be completely separated without error. For example, consider $n = 6$, $k = 3$ (with two points per group), and $p = 5$. Then, the three means lie in two dimensions and each group has two points placed orthogonally to the two-dimensional space containing the means *and* to the

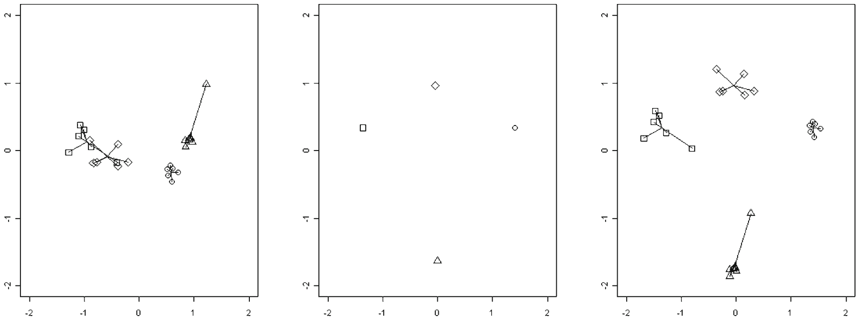

Figure 1. Analysis of the data of Section 3 in, from left to right, (a) the range space, (b) the intersection space, and (c) a four-dimensional combination of these spaces in which the within-group variation is orthogonal to the plane holding the group means. Each type of marker depicts a different slip road, with lines going to the mean values for each slip road.

other within-groups spaces, exhausting the 5 available dimensions. If the within-group points are plotted together with the group means, one has to remember that the within group scatters are in orthogonal spaces. The separability of the groups would occur, however artificially the groups themselves are defined. This observation has implications on the possibilities for finding spurious structure in data-mining investigations that have many variables relative to the number of units.

One use of CVA is as a discrimination method, plotting new samples to see which group mean is nearest (also checking that they do not lie beyond the reasonable range of dispersion, perhaps by supplying confidence circles). In our case a new sample is unlikely to lie totally in one of the orthogonal spaces but we can still check which is the nearest mean and whether or not they lie inside a confidence circle (sphere etc…). This use assumes that generalized Mahalanobis distance is appropriate. Connected with this is the interpretation of D^2. In the classical case D^2 is monotonically related to the probabilistic overlap between two multinormal distributions with different means but the same dispersion. What happens when the distributions are singular depends on whether or not the dispersions of the means are assumed to be in the same space as the within-group dispersion. Solutions in the intersection space (Section 3.2.1), are orthogonal to the within group variation so have no interpretation in terms of probabilistic overlap. The solution of Section 3.2.2 reduces the problem to the classical case by working in the $n - k$ range space of $\mathbf{W}$ or, alternatively, in the range space of $\mathbf{T}$, which includes part of the null space of $\mathbf{W}$, both of which do admit probabilistic interpretation.

In generalizing Mahalanobis distance we have sought solutions using a g-inverse of $\mathbf{W}$. The preceding remarks demonstrate that this does not suffice to identify a unique solution as the MP-metric gives solutions in the range space of **W**, in the intersection space and in the complete (n – 1)-dimensional space which all differ but all are based on the same Moore-Penrose g-inverse. Figures 2, 3 and 4 illustrate the geometry. Note, that in these figures, the range and intersection spaces are drawn as axes that are proxies for multidimensional spaces whose dimensions are indicated by the subscripts given on each "axis". Thus, what look like collinearities are in fact points in these multidimensional spaces.

In all cases the variation in the intersection space gives complete separation between groups but it seems unlikely to represent reality or, indeed, reproducibility. Although mathematically possible, there seems little to support a practical situation where variation between group means is orthogonal to variation within groups. If it were a reality, it would give an infallible way of assigning objects to groups.

Some, but not all, of what has been covered above will have some familiarity. It is not easy to disentangle what is new and what is not. Some things are new, some are seen in a new light and some, although well known, are open to discussion. It is clear that the GCF offers a basic tool that may assuredly be included among the fundamental algebraic decompositions. The GCF has shown how considerations of rank alone uncover the structure of the CVA problems under discussion and enable a unified approach. It has shown how ratio-criteria and model-fitting approaches need to be separately considered and it has shown how model-fitting and choice of metric are linked. It has thrown light on one of the most used generalizations of CVA and shown how it may be improved. It has also uncovered where new work is called for, namely in further study of appropriate metrics and the interpretation of solutions in the intersection space.

Appendix A
The General Canonical Form

In this Appendix **A** and **B** are symmetric definite or p.s.d. matrices. Clearly the null space of **B** may intersect the null space of **A**. Although here the results are put into a generic form, in the main text we shall usually have that **A** will be the within group dispersion matrix **W** while **B** will usually refer to the between group dispersion matrix **B**, itself.

The basic GCF result. Two p.s.d. quadratic forms **A** and **B** may be simultaneously diagonalized by a nonsingular transformation **L** so that:

$$\mathbf{L}'\mathbf{A}\mathbf{L} = \left(\begin{array}{cc|cc} \mathbf{I} & & & \\ & \mathbf{I} & & \\ \hline & & \mathbf{0} & \\ & & & \mathbf{0} \end{array}\right) \quad \text{and} \quad \mathbf{L}'\mathbf{B}\mathbf{L} = \left(\begin{array}{cc|cc} \boldsymbol{\Gamma}^{(1)} & & & \\ & \mathbf{0} & & \\ \hline & & \boldsymbol{\Gamma}^{(0)} & \\ & & & \mathbf{0} \end{array}\right)$$

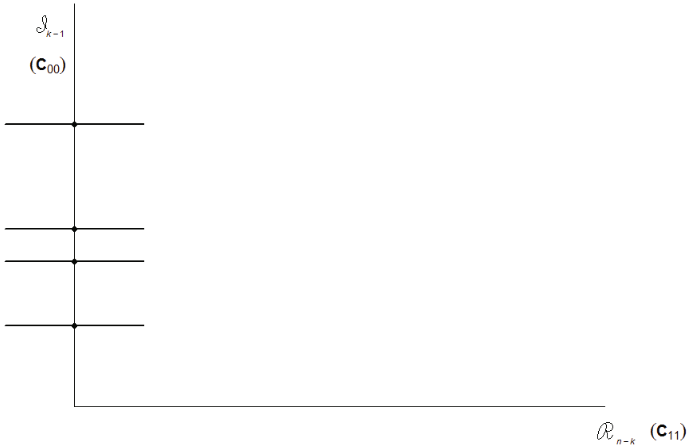

Figure 2. The geometry of the canonical variables given by the GCF (13). Here, *R* refers to the range space of **W** and *I* to the intersection space. The difference between the means $\mathbf{HXL}^{(0)}$ lies entirely in *I* and within-group variation $\mathbf{(I - H)XL}^{(1)}$ entirely in *R*. There is complete separability between groups in *I*.

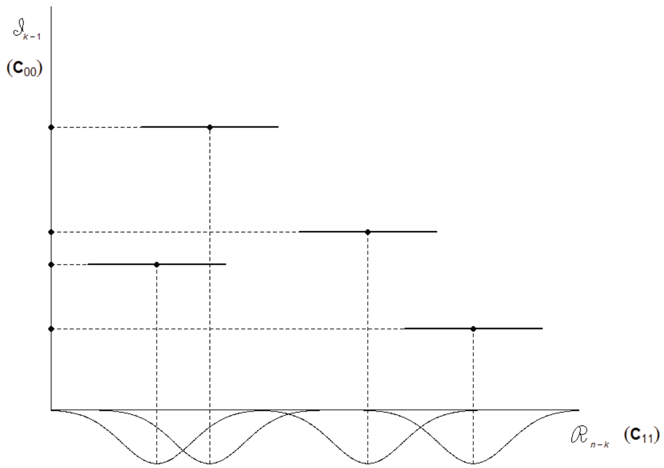

Figure 3. The geometry of the canonical variables in the range space of **T** given by (A5) and Section 2.2.2. Here, *R* refers to the range space of **W** and *I* to the intersection space. The variation is "two-dimensional" being both in *I* and *R*; all within group variation is in *R*. The projection onto *I* gives the separable groups indicated in the intersection space and is the same as in Figure 2. The projection onto *R* is that part of the total variation lying in the range space of **W** and gives the overlapping between-group variation as indicated by the distributional glyphs.

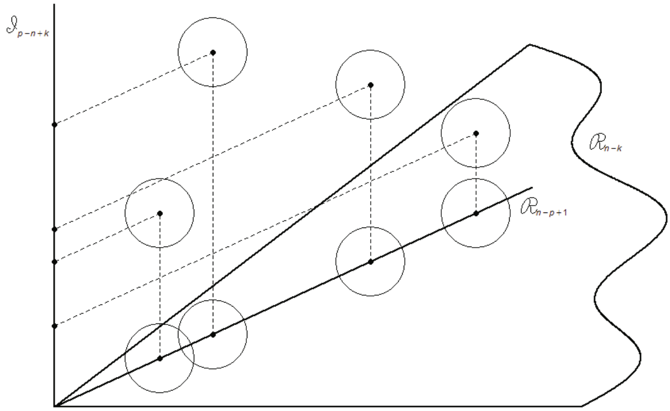

Figure 4. The geometry of the canonical variables in the range space of **T** given by equation (16) of the GCF when $n < p - k < p$. The variation in the n-k dimensional range space R_{n-k} of **W** is shown as a "two-dimensional" region; variation within groups is entirely within this region where there is possible overlap, indicated by notional confidence circles. However, between group variation is partly in I, where there is complete separability, and partly in R_{n-p+k} a subspace of dimension $n - p + k$ in R_{n-k}.

where the horizontal and vertical lines separate the null and range spaces of **A**. The matrices $\mathbf{\Gamma}^{(1)}$ and $\mathbf{\Gamma}^{(0)}$ are diagonal; throughout we use the superscript notation to refer to similar matrices in the range space of $\mathbf{A}^{(1)}$, $\mathbf{B}^{(1)}$ and the null space of $\mathbf{A}^{(0)}$, $\mathbf{B}^{(0}$**,** respectively.

In the following we give the algebraic details of the transformation but the most important result is that **L** is nonsingular. A special case of a more general result given by Albers et al. (2011) shows that the transformation **L** may be written explicitly as follows:

$$\mathbf{L} = (\mathbf{L}^{(1)}, \mathbf{L}^{(0)}), \tag{A1}$$

where

$$\left.\begin{aligned} \mathbf{L}^{(1)} &= \mathbf{U}^{(1)}\mathbf{\Delta}^{-1}\mathbf{V} - \mathbf{U}^{(0)}\mathbf{R}\mathbf{V} \\ \mathbf{L}^{(0)} &= \mathbf{U}^{(0)}\mathbf{Q} \end{aligned}\right\} \tag{A2}$$

$$\text{and } \mathbf{R} = \mathbf{Q}^{(1)}\mathbf{\Gamma}^{(0)-1}\mathbf{Q}'^{(1)}\mathbf{C}_{01}.$$

The inverse transformation is:

$$\mathbf{L}^{-1} = \begin{pmatrix} \mathbf{V}'\mathbf{\Delta}\mathbf{U}'^{(1)} \\ \mathbf{Q}'(\mathbf{R}\mathbf{\Delta}\mathbf{U}'^{(1)} + \mathbf{U}'^{(0)}) \end{pmatrix}. \tag{A3}$$

The orthogonal matrices **U**, **V** and **Q** occurring in (A1), (A2) and (A3) derive from the following spectral decomposition:

$$\mathbf{A} = \mathbf{U}^{(1)}\boldsymbol{\Delta}^2\mathbf{U}'^{(1)}$$

which, in turn, is used to determine the spectral decompositions of matrices derived from the component matrices of **C,** defined below:

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{10} \\ \mathbf{C}_{01} & \mathbf{C}_{00} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Delta}^{-1}\mathbf{U}'^{(1)}\mathbf{B}\mathbf{U}^{(1)}\boldsymbol{\Delta}^{-1} & \boldsymbol{\Delta}^{-1}\mathbf{U}'^{(1)}\mathbf{B}\mathbf{U}^{(0)} \\ \mathbf{U}'^{(0)}\mathbf{B}\mathbf{U}^{(1)}\boldsymbol{\Delta}^{-1} & \mathbf{U}'^{(0)}\mathbf{B}\mathbf{U}^{(0)} \end{pmatrix}$$

with spectral forms:

$$\left.\begin{aligned} \mathbf{C}_{00} &= \mathbf{Q}^{(1)}\boldsymbol{\Gamma}^{(0)}\mathbf{Q}'^{(1)} \\ \mathbf{C}_{11} - \mathbf{C}_{10}(\mathbf{Q}^{(1)}\boldsymbol{\Gamma}^{(0)-1}\mathbf{Q}'^{(1)})\mathbf{C}_{01} &= \mathbf{V}\boldsymbol{\Gamma}^{(1)}\mathbf{V}' \end{aligned}\right\}.$$

In the spectral decomposition of **A**, the complete set of eigenvectors is given by $\mathbf{U} = (\mathbf{U}^{(1)}, \mathbf{U}^{(0)})$ where $\mathbf{U}^{(0)} \perp \mathbf{U}^{(1)}$ but is otherwise arbitrary. Similarly, any null vectors of **Q** and **V** are arbitrary, corresponding to (possibly many) zero values of $\boldsymbol{\Gamma}$ in the spectral decompositions. $\mathbf{Q}^{(1)}$ represents the columns of **Q** corresponding to nonzero values of $\boldsymbol{\Gamma}^{(0)}$, When $\boldsymbol{\Gamma}^{(0)}$ vanishes, we may take **V** = **I**.

To understand the underlying structure of the GFC, we can write (A2) as:

$$\mathbf{L} = \begin{pmatrix} \mathbf{U}^{(1)}\boldsymbol{\Delta}^{-1} & \mathbf{U}^{(0)} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{R} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix},$$

where, (i) the first term induces the transformation of **A** to diagonal 1/0 form, (ii) the second term leaves the 1/0 transformation of **A** unchanged but transforms **B** to two diagonal symmetric blocks with elements $\mathbf{C}_{11}\text{-}\mathbf{C}_{10}\mathbf{C}_{00}^{-1}\mathbf{C}_{01}$ and $\mathbf{C}_{00}$ and (iiii) the third term leaves the 1/0 transformation of **A** unchanged but is an orthogonal matrix that diagonalizes both $\mathbf{C}_{11} - \mathbf{C}_{10}\mathbf{C}_{00}^{-1}\mathbf{C}_{01}$ and $\mathbf{C}_{00}$ to give nonzero diagonal terms $\boldsymbol{\Gamma}^{(1)}$ and $\boldsymbol{\Gamma}^{(0)}$ (but in equation (13) of Section 3.2 we have $\boldsymbol{\Gamma}^{(1)} = \mathbf{0}$ and hence

$$\mathbf{C}_{11} - \mathbf{C}_{10}\mathbf{C}_{00}^{-1}\mathbf{C}_{01} = \mathbf{0}\ ).$$

g-Inverse Property and Metrics. The interest here generalizes the property of Mahalanobis distance that when **W** is nonsingular, the scaling **M'WM** = **I** implies that the Mahalanobis metric $\mathbf{W}^{-1} = \mathbf{MM}'$. When **W** is singular, we are interested in metrics $\mathbf{W}^{-} = \mathbf{LL}'$ based on a g-inverse of **W**. Note that $\mathbf{W}^{-}$ is not itself singular.

Suppose **L'AL** = **J**, a diagonal matrix of units and zeros, it follows immediately that $(\mathbf{L'AL})^2 = \mathbf{L'AL}$ so that with **L** nonsingular **A(LL')A** = **A** and therefore **LL'** is a g-inverse of **A**. Thus, the condition **L'AL** = **J** is sufficient for guaranteeing a g-inverse $\mathbf{W}^-$ of the required form.

Now **L'AL** = **J** is a property of the GCF, so we term, **LL'** the GCF metric. Of course, **LL'** is not a unique g-inverse of **A** as is now shown. From

$$\mathbf{(L'AL)J(L'AL) = J}$$

we have that:

$$\mathbf{A(LJL')A} = (\mathbf{L}')^{-1}\,\mathbf{J}\,(\mathbf{L})^{-1} = (\mathbf{L}')^{-1}\,(\mathbf{L'AL})\,\mathbf{L}^{-1} = \mathbf{A},$$

showing that **LJL'** is also a g-inverse of **A** but it is singular, unlike **LL'** itself.

An important special case is when **L** arises from the first step used in deriving the GCF. We use $\mathbf{L}^*$ to denote that in this section we are replacing **L** by the definitions given in (A4):

$$\mathbf{L}^{*(1)} = \mathbf{U}^{(1)}\boldsymbol{\Delta}^{-1},\ \ \mathbf{L}^{*(0)} = \mathbf{U}^{(0)}. \qquad \text{(A4)}$$

As before, $\mathbf{L}^{*}{}'\mathbf{A}\ \mathbf{L}^{*} = \mathbf{J}$ and so $\mathbf{L}^{*}\mathbf{L}^{*}{}'$ continues to satisfy the g-inverse property. Indeed, $\mathbf{L}^{*}\mathbf{L}^{*}{}' = \mathbf{A}^{+}$, the Moore Penrose inverse of **A,** so we term $\mathbf{L}^{*}\mathbf{L}^{*}{}'$ the MP-metric.

A general symmetric g-inverse **A*** of **A** may be written (see Rao1967 and Gower 1976) in terms of some particular g-inverse **LL'**, to give for arbitrary **P**:

$$\mathbf{A^* = LL' + (I - LL'A)P + P'(I - ALL')}. \qquad \text{(A5)}$$

Thus, **A*** has arbitrary components and in general the distances with which we are concerned depend on the choice of **P**.

Choosing the MP-metric as our basic g-inverse, we have from (A4) that $\mathbf{LL'A} = \mathbf{U}^{(1)}\mathbf{U}'^{(1)}$ and substituting into (A5) gives:

$$\mathbf{A^*} = (\mathbf{A}^{+}) + (\mathbf{I} - \mathbf{U}^{(1)}\mathbf{U}'^{(1)})\mathbf{P} + \mathbf{P}'(\mathbf{I} - \mathbf{U}^{(1)}\mathbf{U}'^{(1)}) \qquad \text{(A6)}$$

$$= (\mathbf{A}^{+}) + (\mathbf{U}^{(0)}\mathbf{U}'^{(0)})\mathbf{P} + \mathbf{P}'(\mathbf{U}^{(0)}\mathbf{U}'^{(0)}). \qquad \text{(A7)}$$

The choice of **P** in the g-inverse governs the choice of metric (e.g. (A2) or (A4)). When **P** = 0 we have the Moore-Penrose inverse (A4); also **P** = **-½I** delivers the g-inverse **LJL'**. The particular case when **L** is given by (A2), associated with the GCF-metric, arises from setting **P** = $\mathbf{U}^{(0)}\mathbf{R}\left(\tfrac{1}{2}\mathbf{R}'\mathbf{U}'^{(0)} - \boldsymbol{\Delta}^{-1}\mathbf{U}'^{(1)}\right)$ in (A7).

We are concerned with Mahalanobis- type distances derived from inner products of the form **TX(A*)X'T'** for various choices of **T**. It is a reasonable requirement that these distances be independent of the partially arbitrary nature of $\mathbf{U}^{(0)}$. When **P**= 0, we have from (A4) that:

$$\mathbf{LL'} = \mathbf{A}^{+} = \mathbf{U}^{(1)}\boldsymbol{\Delta}^{-2}\mathbf{U}'^{(1)} + \mathbf{U}^{(0)}\mathbf{U}'^{(0)} = \mathbf{U}^{(1)}\boldsymbol{\Delta}^{-2}\mathbf{U}'^{(1)} + \mathbf{I} - \mathbf{U}^{(1)}\mathbf{U}'^{(1)}, \quad \text{(A8)}$$

which depends only on $\mathbf{U}^{(1)}$, and it follows that (A7) is independent of $\mathbf{U}^{(0)}$ as are derived Mahalanobis distances. This property does not hold when $\mathbf{L}$ is given by (A2) as used in the GCF-metric, when (A7) becomes:

$$\mathbf{A}^{*} = (\mathbf{A}^{+}) + (\mathbf{U}^{(0)}\mathbf{RR}'\mathbf{U}'^{(0)} - \mathbf{U}^{(0)}\mathbf{R}\Delta^{-1}\mathbf{U}'^{(1)} - \mathbf{U}^{(1)}\Delta^{-1}\mathbf{R}'\mathbf{U}'^{(0)}). \quad \text{(A9)}$$

A sufficient property for independence from $\mathbf{U}^{(0)}$ is that $\mathbf{TXU}^{(0)} = 0$.

Appendix B
Calculating Spectral Structure for Large Matrices (p >> n)

The notation in this section **X** refers to general matrices and the spectral structure of their inner products. In our applications **X** may be the same as in the main text and Appendix **A**, or it may be **(I – H)X** of the main text.

Assuming that **X** is column-centred and so has rank $n - 1$, the usual spectral decomposition required for large values of p is:

$$\left({}_p\mathbf{X}'_n\mathbf{X}_p\right)\mathbf{U}_p = \mathbf{U}\begin{pmatrix}\boldsymbol{\Delta}^2_{n-1} & \\ & \mathbf{0}\end{pmatrix}$$

from which the first n-1 columns give: ${}_p(\mathbf{X}'\mathbf{X})_p\mathbf{U}^{(1)}_{n-1} = {}_p\mathbf{U}^{(1)}_{n-1}\boldsymbol{\Delta}^2_{n-1}$ and the final $p - n + 1$ columns give ${}_p(\mathbf{X}'\mathbf{X})_p\mathbf{U}^{(0)}_{p-n+1} = \mathbf{0}$.

For large values of p computing this decomposition is a major problem. A smaller eigenvalue problem is to solve:

$$\left({}_n\mathbf{X}_p\mathbf{X}'\right){}_n\mathbf{X}_p\mathbf{U}^{(1)}_{\mathbf{n}-1} = {}_n\mathbf{X}_p\mathbf{U}^{(1)}_{\mathbf{n}-1}\boldsymbol{\Delta}^2_{n-1}$$

with the same non-zero eigenvalues as for **X'X** and eigenvectors ${}_n\mathbf{Q}_{n\text{-}1} = \mathbf{XU}^{(1)}$ scaled so that ${}_{n-1}\mathbf{Q}'_n\mathbf{Q}_{n-1} = \boldsymbol{\Delta}^2$.

Given the eigenvectors **Q** and **Δ** derived and scaled as above, we may obtain $\mathbf{U}^{(1)}$ as follows:

$$\mathbf{X}'_{\mathrm{n}}\mathbf{Q}_{\mathrm{n}-1}\boldsymbol{\Delta}^{-1}_{\mathrm{n}-1} = \mathbf{X}'_{\mathrm{n}}\mathbf{X}_{\mathrm{p}}\mathbf{U}^{(1)}_{\mathrm{n}-1}\boldsymbol{\Delta}^{-1}_{\mathrm{n}-1} = {}_{\mathrm{p}}\mathbf{U}^{(1)}_{\mathrm{n}-1}\boldsymbol{\Delta}_{\mathrm{n}-1}$$

Hence:

$${}_{\mathrm{p}}\mathbf{U}^{(1)}_{\mathrm{n}-1} = {}_{\mathrm{p}}\mathbf{X}'_{\mathrm{n}}\mathbf{Q}_{\mathrm{n}-1}\boldsymbol{\Delta}^{-2}_{\mathrm{n}-1}.$$

We may obtain $\mathbf{U}^{(0)}$ as any p–n+1 columns of $\mathbf{I}_p$ - $\mathbf{U}^{(1)}\mathbf{U}'^{(1)}$. However, it is computationally more convenient to eliminate the common null space altogether and this is easily done by referring **X** to its principal axes, i.e. replacing **X** by $\mathbf{XU}^{(1)} = {}_n\mathbf{Q}_{n\text{-}1}$. Working with **Q** eliminates all the zero blocks in the final rows and columns of (12) and reduces the problem to the case $n - k < p \leq n - 1$. In particular, all the $k - 1$ null vectors ${}_{n\text{-}1}\mathbf{U}^{(0)}_{k\text{-}1}$ in the intersection spaces of **T** and **W** are found from the spectral decomposition of **W** derived from ${}_n\mathbf{Q}_{n\text{-}1}$.

References

ALBERS, C.J., CRITCHLEY, F., and GOWER, J.C. (2011a), "Quadratic Minimisation Problems in Statistics", *Journal of Multivariate Analysis*, *102*, 698–713.

ALBERS, C.J., CRITCHLEY, F., and GOWER, J.C. (2011b), "Applications of Quadratic Minimisation Problems in Statistics", *Journal of Multivariate Analysis*, *102*, 714–722.

CHEN, L., LIAO, H., KO, M., LIN, J., and YU, G. (2000), "A New LDA-Based Face Recognition System Which Can Solve the Small Sample Size Problem", *Pattern Recognition*, *33*, 1713–1726.

ECKART, C., and YOUNG, G. (1936), "The Approximation of One Matrix by Another of Lower Rank", *Psychometrika, 1*, 211 – 218.

GOWER, J.C., and ALBERS, C.J. (2011), "Between Group Metrics", *Journal of Classification*. *28*, 315–326.

GOWER, J.C. (1976), "Growth-Free Canonical Variates and Generalised Inverses", *Bulletin of the Geological Institute of the University of Uppsala 7*, 1–10.

GOWER, J.C. (1998), "The Role of Constraints in Determining Optimal Scores", *Statistics in Medicine*, *17*, 2709–2721.

HEALY, M.J.R., and GOLDSTEIN, H. (1976), "An Approach to the Scaling of Categorised Attributes", *Biometrika*, *63*, 219–229.

KRZANOWSKI, W.J., JONATHAN, P., MCCARTHY, W.V., and THOMAS, M.R. (1995), "Discriminant Analysis with Singular Covariance Matrices: Methods and Applications to Spectroscopic Data", *Journal of the Royal Statistical Society, Series C (Applied Statistics*), *44*, 101–115.

LUBBE-GARDNER, D.L., LE ROUX, N.J., and GOWER, J.C. (2008), "Measures of Fit in Principal Component and Canonical Variate Analysis", *Journal of Applied Statistics, 35***,** 947–965.

MARDIA, K.V. (1977), "Mahalanobis Distances and Angles", in *Multivariate Analysis (IV)*, ed. P.R. Krishnaiah, Amsterdam: North-Holland, pp. 495–511.

QUEEN, C.M., WRIGHT, B., and ALBERS, C.J. (2007), "Eliciting a Directed Acyclic Graph for a Multivariate Time Series of Vehicle Counts in a Traffic Network", *Australian and New Zealand Journal of Statistics, 49(3),* 1–19.

QUEEN, C.M., and ALBERS, C.J. (2009), "Intervention and Causality: Forecasting Traffic Flows Using a Dynamic Bayesian Network", *Journal of the American Statistical Association*, *104(486),* 669–681.

RAO, C.R. (1949), "Representation of *p*-Dimensional Data in Lower Dimensions", in *Anthropometric Survey of the United Provinces: A Statistical Study, 1941*, by Mahalanobis, P.C., Majumdar, D.N. and Rao, C.R., in *Sankhya*, *The Indian Journal of Statistics, 9,* 90–324.

RAO, C.R. (1967), "Calculus of Generalised Inverses of Matrices: Part 1 – General Theory", *Sankhya, Series A, 29*, 317–342.

RAO, C.R., and YANAI, H. (1979), "General Definition and Decomposition of Projectors and Some Applications to Statistical Problems", *Journal of Statistical Planning and Inference 3,* 1–17.

YE, J. (2005), "Characterization of a Family of Algorithms for Generalized Discriminant Analysis on Undersampled Problems", *Journal of Machine Learning Research, 6,* 483–502.

YE, J., and XIONG, T. (2006), "Computational and Theoretical Analysis of Null Space and Orthogonal Linear Discriminant Analysis", *Journal of Machine Learning Research*, *7*, 1183–1204.